

Supplementary Material for:

Fluostatins produced by the heterologous expression of a TAR reassembled environment DNA derived Type II PKS gene cluster.

Zhiyang Feng, Jeff H. Kim, and Sean F. Brady*

Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University,
New York, NY 10021, Fax: (212) 327-8281, E-mail: sbrady@rockefeller.edu

RECEIVED DATE (automatically inserted by publisher); Add Author E-mail Address Here

Clone identification and TAR reconstruction:

The eDNA mega-library was constructed following a previously described protocol using pWEB-TNC (Epicentre).¹ AB649 was originally identified and recovered from this library using minimal PKS specific degenerate primers (dp:KS: TTCGGSGGITCCAGWSIGCSATG and dp:ACP: TCSAKSAGSGCSAISGASTCGTAICC).^{2,3}

The two overlapping clones used in this study (AB649 and AB1850) were recombined into one intact type II PKS pathway using TAR and the yeast/*E. coli*/Streptomyces shuttle capture vector pTARa.^{4, 5} AB649 and AB1850 specific homology arms were generated by PCR from AB649 and AB1850 using the following two primer pairs:

649T7F(*Nhe*I): GCGCGCTAGCGCGGTGGTTGTTGGTGAGAAGC

649T71850T7R(*Pme*I): GTCGGGGTGATGGCGTGGTAAACCCGACTACACCTCCGGTGAG

and

1850T7649T7F(*Pme*I): CTCACCGGAAGGTGTAGTCGGTTAAACCGAACACGCCATACCCCCGAC

1850T7R(*Sph*I): GCGCGCATGCCGAATCCGAATCCCTTCGTCG.

These two PCR products served as templates in a second round of PCR with primers 649T7F(*Nhe*I) and 1850T7R(*Sph*I). The resulting PCR product was doubly digested with *Sph*I and *Nhe*I and ligated to a similarly digested pTARa capture vector. Two hundred ng of *Dra*I digested AB649 and AB1850 were mixed with 100 ng of *Pme*I digested capture vector and the mixture was then added to 200 µL of *Saccharomyces cerevisiae* spheroplasts. This spheroplast DNA mixture was then added directly to 7 mL of top agar (1 M sorbitol, 1.92 g/L SC uracil dropout supplement, 6.7 g/L yeast nitrogen base, 2% glucose, 2.5% agar). The top agar containing spheroplasts was overlayed onto SC dropout agar and the plates incubated at 30°C. Yeast colonies were picked after 72 hours and grown overnight at 30°C in YPD. Vector DNA extracted from these cultures (ChargeSwitch™, Invitrogen) was transformed directly into EPI300 *E. coli* (Epicentre) and the BAC DNA obtained from EPI300 was used for downstream analysis. The TAR product as well as each individual eDNA clone was fully sequenced using 454 pyrosequencing. All clones were introduced into *Streptomyces albus* by conjugation using *E. coli* S17.⁶ Prior to conjugation individual eDNA clones were retrofitted with the *Dra*I fragment from pOJ436.^{2,6}

(1) Brady, S. F. Nat. Protoc. 2007, 2, 1297-1305.

(2) King, R. W.; Bauer, J. D.; Brady, S. F. Angew. Chem. Int. Edit. 2009, 48, 6257-6261.

- (3) Seow, K. T.; Meurer, G.; Gerlitz, M.; WendtPienkowski, E.; Hutchinson, C. R.; Davies, J. J. *Bacteriol.* 1997, 179, 7360-7368.
 (4) Gibson, D. G, et. al. *Science* 2008, 319, 1215-1220.
 (5) Kim, J.; Feng, Z.; Bauer, J. D.; D., K.; Calle, P. Y.; Brady, S. F. *Biopolymers* 2010, in press.
 (6) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces genetics*; The John Innes Foundation: Norwich, United Kingdom, 2000.

Isolation and purification of rabelomycin and dehydrorabelomycin:

Crude HP-20 eluant containing rabelomycin and dehydrorabelomycin was initially fractioned by silica gel flash chromatography using a CHCl₃:MeOH step gradient. Rabelomycin and dehydrorabelomycin eluted from this column with 100:0 and 90:10 CHCl₃:MeOH, respectively. Each metabolite was then purified by reversed phase (XBridgeTM C18, 150 x 10 mm, 5 µm) HPLC (7ml/min) using a linear gradient of 90:10 H₂O:MeOH (containing 0.1% formic acid) to 100% MeOH (containing 0.1% formic acid) over 30 min (rabelomycin 18.2 min, dehydrorabelomycin 27.8 min).

Rabelomycin (**1**) HRMS-ESI (*m/z*): [M+Na]⁺: calcd for C₁₉H₁₄O₆Na 361.0688; found 361.0684. ¹H NMR (600 MHz, CDCl₃) δ 12.29 (s, 1H), 11.64 (s, 1H), 7.66 (t, *J* = 7.4 Hz) 7.65 (d, *J* = 7.4 Hz), 7.25 (d, *J* = 7.4 Hz, 1H), 7.00 (s, 1H), 3.07 (s, 2H), 3.03 (d, *J* = 15.1 Hz, 1H), 2.94 (d, *J* = 15.1 Hz, 1H), 1.48 (s, 3H).

Dehydrorabelomycin (**2**) ¹H NMR (600 MHz, CDCl₃) δ 12.10 (s, 1H), 11.74 (s, 1H), 10.28 (s, 1H), 7.83 (d, *J* = 7.5 Hz, 1H), 7.70 (dd, *J* = 8.5, 7.5 Hz, 1H), 7.64 (s, 1H), 7.32 (d, *J* = 8.5 Hz, 1H), 7.07 (s, 1H), 6.93 (s, 1H), 2.42 (s, 3H).

Fluostatin C (**3**) HRMS-ESI (*m/z*): [M-H]⁻: calcd for C₁₈H₁₁O₆ 323.0556; found 323.0538. ¹H NMR (600 MHz, *d*-Acetone) δ 7.53 (s, 1H), 7.30 (dd, *J* = 8.1, 7.2 Hz, 1H), 7.21 (d, *J* = 7.2 Hz, 1H), 7.10 (d, *J* = 8.1 Hz, 1H), 6.02 (d, *J* = 2.3 Hz, 1H), 3.87 (d, *J* = 2.3 Hz, 1H), 1.56 (s, 3H).

¹³C NMR labeling results:

Table S1. ¹³C integrations observed for fluostatin G isolated from [1-¹³C] acetate fed and unlabeled cultures. Integrations are normalized to the signal observed for C-12.

Position of the carbon	C1	C2	C3	C4	C4a	C5	C6	C6a	C6b
unlabeled fluostatinG	0.97	0.92	0.57	0.31	0.29	0.83	0.38	0.30	0.31
labeled fluostatinG	35	0.94	18.03	0.39	12.57	0.95	0.48	13.22	0.42
Ratio	36	1.0	32	1.3	43	1.1	1.3	44	1.4

Position of the carbon	C7	C8	C9	C10	C10a	C11	C11a	C11b	C12
unlabeled fluostatinG	0.44	0.83	0.82	0.71	0.35	0.27	0.26	0.30	1
labeled fluostatinG	16.5	0.93	30.09	0.82	10.10	0.38	9.09	0.32	1
Ratio	38	1.1	37	1.2	29	1.4	35	1.1	1

Table S2. ^1H - ^{13}C HMBC data for fluostatin F

Table S3. ^1H - ^{13}C HMBC data for fluostatin G

Table S4. ^1H - ^{13}C HMBC data for fluostatin H

Table S5. Gene table for AB649/1850.

Gene	Clone	Size	Deduced function	% I/S	Origin	Accession no.
<i>orf1</i>	649	318	beta-lactamase domain protein	67/76	<i>Micromonospora aurantiaca</i>	ZP_06221109.1
<i>orf2</i>	649	294	beta-lactamase	68/80	<i>Micromonospora</i> sp.	ZP_06398714.1
<i>orf3</i>	649	318	transcriptional regulator	61/69	<i>Streptomyces griseoflavus</i>	ZP_05543482.1
<i>orf4</i>	649	557	carboxyl transferase	79/86	<i>S. hygroscopicus</i>	ZP_05514554.1
<i>orf5</i>	649	4862	bifunctional monoxygenase/dehydratase	68/74	<i>S. venezuelae</i>	AAV52248.1
<i>orf6</i>	649	501	oxygenase reductase-like protein	60/71	<i>S. murayamaensis</i>	AAO65351.1
<i>orf7</i>	649	316	aromatase	69/79	<i>S. antibioticus</i>	CAG14969.1
<i>orf8</i>	649	260	ketoreductase	80/88	<i>S. antibioticus</i>	CAG14968.1
<i>orf9</i>	649	91	acyl carrier protein	56/70	<i>S. cyanogenus</i>	AAD13538.1
<i>orf10</i>	649	420	chain length determinant	73/82	<i>S. venezuelae</i>	AAB36563.1
<i>orf11</i>	649	426	putative ketoacyl synthase	84/90	<i>S. fradiae</i>	CAA60569.1
<i>orf12</i>	649	109	putative polyketide cyclase	83/85	<i>S. ambofaciens</i>	AAR30153.1
<i>orf13</i>	649	578	acetyl/proionyl CoA carboxylase	68/78	<i>S. pristinaespiralis</i>	ZP_05009349.1
<i>orf14</i>	649	228	regulator	72/85	<i>S. venezuelae</i>	AAB36584.2
<i>orf15</i>	649	133	hypothetical protein SBI_07207	54/63	<i>S. bingchenggensis</i>	ADI10327.1
<i>orf16</i>	649	496	oxygenase-like protein	65/77	<i>S. murayamaensis</i>	AAO65343.1
<i>orf17</i>	649	237	anthrone oxidase-like protein	58/75	<i>S. murayamaensis</i>	AAO65344.1
<i>orf18</i>	649	179	hypothetical protein	28/46	<i>Rhizobium etli</i>	YP_472332.1
<i>orf19</i>	649	364	hypothetical protein	36/50	<i>Xylanimonas cellulosilytica</i>	YP_003326777.1
<i>orf20</i>	649	253	hypothetical protein RHA1_ro04727	61/75	<i>Rhodococcus jostii</i>	YP_704670.1
<i>orf21</i>	649	491	putative oxygenase	58/71	<i>S. ambofaciens</i>	CAK51002.1
<i>orf22</i>	649	407	monooxygenase FAD-binding protein	58/74	<i>Stackebrandtia nassauensis</i>	YP_003513318.1
<i>orf23</i>	649	335	putative O-methyltransferase	57/70	<i>S. ambofaciens</i>	AAR30145.1
<i>orf24</i>	649	259	hypothetical protein Strop_2177	63/74	<i>Salinispora tropica</i>	YP_001159006.1
<i>orf25</i>	649	253	3-hydroxyacyl-CoA dehydrogenase	61/74	<i>Frankia alni</i>	YP_713325.1
<i>orf26</i>	649	323	Alcohol dehydrogenase zinc- binding domain protein	67/80	<i>Thermomonospora curvata</i>	YP_003301251.1
<i>orf27</i>	649	344	3-oxoacyl-(acyl carrier protein) synthase III	42/61	<i>Symbiobacterium thermophilum</i>	YP_075277.1
<i>orf28</i>	649	241	short-chain dehydrogenase/reductase SDR	63/77	<i>Sa. tropica</i>	YP_001159019.1
<i>orf29</i>	649	485	SaqE, monooxygenase	55/67	<i>Micromonospora</i> sp.	ACP19351.1
<i>orf30</i>	649/ 1850	330	streptomycin biosynthesis operon regulator	48/62	<i>S.griseoflavus</i>	ZP_05537706.1
<i>orf31</i>	649/ 1850	400	nitric-oxide synthase	53/68	<i>Amycolatopsis mediterranei</i>	ADJ45931.1
<i>orf32</i>	1850	279	nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	39/57	<i>Ktedonobacter racemifer</i>	ZP_06971257.1
<i>orf33</i>	1850	133	GCN5-related N-acetyltransferase	68/75	<i>Sa. tropica</i>	YP_001159036.1
<i>orf34</i>	1850	424	adenylosuccinate lyase	71/82	<i>Sa. tropica</i>	YP_001159035.1
<i>orf35</i>	1850	498	amidase	60/69	<i>Sa. tropica</i>	YP_001159034.1
<i>orf36</i>	1850	501	glutamine synthetase, catalytic region	64/74	<i>Sa. tropica</i>	YP_001159033.1
<i>orf37</i>	1850	119	putative ferredoxin	75/85	<i>S. ambofaciens</i>	CAI78079.1
<i>orf38</i>	1850	628	hypothetical protein	59/70	<i>S. ambofaciens</i>	CAK51011.1
<i>orf39</i>	1850	229	response regulator	68/83	<i>Saccharomonospora viridis</i>	YP_003132178.1
<i>orf40</i>	1850	87	Hypothetical protein CBG08669	44/48	<i>Caenorhabditis briggsae</i>	XP_002636369.1

Table S5. Gene Table for AB649/1850 (continue).

Genes	Clone	Size	Deduced function	% I/S	Origin	Accession no.
<i>orf41</i>	1850	713	PAS/PAC sensor signal transduction histidine kinase	47/59	<i>Actinosynnema mirum</i>	YP_003100166.1
<i>orf42</i>	1850	146	ShdA	37/45	<i>Salmonella enterica</i>	ZP_04657363.1
<i>orf43</i>	1850	302	antitoxin component	32/51	<i>Streptomyces</i> sp.	ZP_06709520.1
<i>orf44</i>	1850	63	putative regulatory protein (partial)	55/63	<i>Frankia alni</i>	YP_715660.1
<i>orf45</i>	1850	66	putative regulatory protein (partial)	55/64	<i>Frankia alni</i>	YP_715660.1
<i>orf46</i>	1850	180	conserved hypothetical protein	33/45	<i>Corynebacterium genitalium</i>	ZP_05707388.1
<i>orf47</i>	1850	372	geranylgeranyl reductase	60/70	<i>Micromonospora aurantiaca</i>	ZP_06215683.1
<i>orf48</i>	1850	181	galactoside O-acetyltransferase	69/77	<i>Sorangium cellulosum</i>	YP_001612024.1
<i>orf49</i>	1850	188	transcriptional regulator, TetR family	51/66	<i>S. roseum</i>	YP_003343267.1
<i>orf50</i>	1850	402	major facilitator superfamily MFS_1 (transporter)	48/63	<i>Cellulomonas flavigena</i> .	YP_003637774.1
<i>orf51</i>	1850	148	unnamed protein product	34/43	<i>Tetraodon nigroviridis</i>	CAG12256.1
<i>orf52</i>	1850	107	AGR369Wp	37/48	<i>Ashbya gossypii</i>	NP_987035.1
<i>orf53</i>	1850	146	hypothetical protein MCAG_03253	48/63	<i>Micromonospora</i> sp.	ZP_04606996.1
<i>orf54</i>	1850	179	intracellular protease, Pfpl family	71/82	<i>Micromonospora</i> sp	ZP_06399106.1
<i>orf55</i>	1850	428	integral membrane protein	49/66	<i>S. hygroscopicus</i>	ZP_05518346.1
<i>orf56</i>	1850	129	NUDIX hydrolase	55/69	<i>Micromonospora</i> sp.	ZP_04607002.1
<i>orf57</i>	1850	312	Ku protein	79/90	<i>Micromonospora</i> sp.	ZP_06399116.1
<i>orf58</i>	1850	315	DNA polymerase LigD ligase region	63/76	<i>Micromonospora</i> sp.	ZP_04607013.1
<i>orf59</i>	1850	302	DNA polymerase LigD, polymerase domain protein	70/83	<i>Micromonospora aurantiaca</i>	ZP_06217829.1
<i>orf60</i>	1850	162	hypothetical protein MCAG_04698	35/47	<i>Micromonospora</i> sp.	ZP_04608441.1
<i>orf61</i>	1850	113	transcriptional regulator, PadR-like family	59/72	<i>Conexibacter woeselii</i>	YP_003393775.1
<i>orf62</i>	1850	141	hypothetical protein Sthe_2782	46/62	<i>Sphaerobacter thermophilus</i>	YP_003321017.1
<i>orf63</i>	1850	553	helicase domain protein	92/96	<i>Micromonospora aurantiaca</i>	ZP_06217838.1
<i>orf64</i>	1850	799	hypothetical protein ML5DRAFT_4875	64/74	<i>Micromonospora</i> sp.	ZP_06400988.1
<i>orf65</i>	1850	77	succinate dehydrogenase and fumarate reductase iron-sulfur protein	43/55	<i>Micromonospora</i> sp.	ZP_04605671.1
<i>orf66</i>	1850	218	HAD-superfamily hydrolase, subfamily IA, variant 1	66/76	<i>Micromonospora</i> sp.	ZP_06400989.1

The closest relatives of the opening reading frames that lie upstream of ORF3 and downstream of ORF52 in the TAR product are, in almost all cases, found in *Micromonospora* spp. For open reading frames that lie between these two ORFs, the closest relatives are found in a diverse group of bacteria, many of which are *Streptomyces* spp. The fluostatin gene cluster likely resides within this island of mostly *Streptomyces* related opening reading frames.